

Tension Bounds for Information Complexity

Manoj M. Prabhakaran*

Vinod M. Prabhakaran[†]

August 28, 2014

Abstract

The main contribution of this work is to relate information complexity to “tension” [PP14] – an information-theoretic quantity defined with no reference to protocols – and to illustrate that it allows deriving strong lower-bounds on information complexity. In particular, we use a very special case of this connection to give a quantitatively tighter connection between information complexity and discrepancy than the one in [BW12] (albeit, restricted to independent inputs). Further, as tension is in fact a multi-dimensional notion, it enables us to bound the 2-dimensional region that represents the *trade-off* between the amounts of communication in the two directions, in a 2-party protocol.

This work is also intended to highlight tension as a fundamental measure of correlation between a pair of random variables, with rich connections to a variety of questions in computer science and information theory.

*Department of Computer Science, University of Illinois, Urbana-Champaign. mmp@illinois.edu.

[†]School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai, India. vinodmp@tifr.res.in

1 Introduction

Communication complexity, since the seminal work of Yao [Yao79], has been a central question in theoretical computer science. Many of the recent advances in this area have centred around the notion of information complexity, which measures the *amount of information* about the inputs – rather than the *number of bits* – that should be present in a protocol’s transcript, if it should compute a function (somewhat) correctly.

The main contribution of this work is to relate information complexity to “tension” [PP14] – an information-theoretic quantity defined with no reference to protocols – and to illustrate that it allows deriving strong bounds on information complexity. In particular, we use a very special case of this connection to give a quantitatively tighter connection between information complexity and discrepancy than the one in [BW12] (albeit, restricted to independent inputs). Further, as tension is in fact a multi-dimensional notion, it enables us to bound the 2-dimensional region that represents the *trade-off* between the amounts of communication in the two directions, in a 2-party protocol.

This work is also intended to highlight tension as a fundamental measure of correlation between a pair of random variables, with rich connections to a variety of questions in computer science and information theory. Tension is intimately related to the notion of *common information* developed in highly influential works in the information theory literature from the 70’s [GK73, Wyn75]. Tension has proven useful in deriving state-of-the-art bounds on “cryptographic complexity” (i.e., number of instances of, say, oblivious transfer needed per instance of securely computing a function) [PP14] and communication complexity of information-theoretically secure multiparty computation [DPP14]. However, currently we have few tools to compute (or bound) tension. We leave it as an important problem to understand tension in general as well as for specific random variables.

What is Tension? Tension of a pair of correlated random variables $(A; B)$ captures “non-trivial” correlation between them: i.e., the extent to which correlation *cannot* be captured by a common random variable that can be associated with both A and B . The question of how well correlation *can* be captured by a random variable is formulated in terms of “common information.” Two different notions of common information were developed in the 70’s, $CI_{\text{GK}}(A; B)$ by Gács-Körner [GK73], and $CI_{\text{Wyn}}(A; B)$ by Wyner [Wyn75], with operational meanings related to certain natural information theoretic problems. (See [Appendix A](#) for more details.) One can define corresponding notions of tension as the gap between mutual information (which accounts for all the correlation, but may not correspond to a common random variable) and common information. More precisely, one can define the non-negative tension quantities $T_{\text{GK}}(A; B) = I(A; B) - CI_{\text{GK}}(A; B)$ and $T_{\text{Wyn}}(A; B) = CI_{\text{Wyn}}(A; B) - I(A; B)$. These notions of tension were identified in [PP14] as special cases of a unified 3-dimensional notion of *tension region*.

In [PP14], an operational meaning was attached to tension region in terms of a communication problem, and also it was shown that a *secure* 2-party protocol for sampling correlated random variables with “high tension”¹ will need a large number of instances of oblivious transfer. In [Appendix A](#), we summarize some of the basic properties of the tension region, as developed in [PP14].

We lower bound the information complexity of a function f in terms of how different the tension regions of $(X; Y)$ and $(X, Z; Y, Z)$ are, where $Z = f(X, Y)$ (or rather, $\Pr[Z = f(X, Y)] \geq \frac{1}{2} + \epsilon$). In particular, when the inputs $(X; Y)$ are independent of each other (so that their tension is zero, and hence contains the origin), the *information complexity region* is shown to lie inside the tension region of $(X, Z; Y, Z)$. (An information complexity region farther from the origin corresponds to a higher lower-bound on information complexity.) Note that even though Z may be a single bit, the difference between the tension regions of $(X; Y)$ and $(XZ; YZ)$ could be quite large – as we illustrate by the connection with discrepancy.

1.1 Overview of Results and Techniques

Our contributions are in two parts:

¹Informally, the farther the tension region is from the origin, the higher the tension, along different dimensions.

1. We show that information complexity can be lower-bounded using tension – a fundamental quantity defined with no reference to protocols.
2. We illustrate the potential of this approach for yielding strong lower-bounds, by obtaining an improved lower-bound on information complexity in terms of discrepancy.

Below, we shall elaborate on these contributions further. We point out that our model and results are, in some ways, more general than prior work:

- In developing the connection between information complexity and tension (as well as between information complexity and communication complexity), we work with a “bigger picture” that considers 2-dimensional notions of these quantities. We remark that even if we are interested only in bounding communication complexity and information complexity (corresponding to 1-dimensional regions), using bounds in terms of the 2-dimensional region can yield potentially stronger lower-bounds.
- Our results hold for randomized functions, with asymmetric outputs.
- A minor difference is that in our communication model, we allow for the possibility that the transcript (i.e., the concatenation of all the messages sent during the protocol in either direction) may not be “parsable” into individual messages by an outsider, though each party, with its input can parse it. (See [Footnote 4](#).)

We propose, as a direction for further study, that various results on information complexity which led to advances in communication complexity can be rederived for tension, thereby providing alternate (and hopefully simpler) proofs to these results. Also, we leave it as an open problem to exploit the full power of the tension bounds: currently, there are few techniques to map out the full 3-dimensional tension region of a pair of random variables.

Tension, Information Complexity and Communication Complexity

The basic idea behind lower-bounding information complexity by tension is, in fact, easy to see. Consider a protocol in which, for simplicity, the two parties are given independent inputs X, Y , exchange messages to generate a transcript M , and produces a common output Z . Since X, Y were independent of each other, we know that (X, Z) and (Y, Z) should continue to be independent conditioned on the transcript, M ; i.e., $(X, Z) - M - (Y, Z)$. One can see that the information cost of this protocol $I(X; M|Y) + I(Y; M|X)$ can be lower bounded by $I(XZ; M|YZ) + I(YZ; M|XZ)$, which in turn can be lower bounded by $\inf_{Q: XZ-Q-YZ} I(XZ; Q|YZ) + I(YZ; Q|XZ)$ (i.e., without requiring that Q is the transcript of a protocol that outputs Z , but only that $XZ - Q - YZ$). The latter quantity is exactly the Wyner-Tension, $T_{\text{Wyn}}(XZ; YZ)$. When (X, Y) are not independent, this lower-bound changes to $T_{\text{Wyn}}(XZ; YZ) - T_{\text{Wyn}}(X; Y)$. Jumping ahead, we mention that we can extend this basic lower-bound to a more general one, where we also consider Q such that the condition $XZ - Q - YZ$ is replaced by $I(XZ; YZ|Q) \leq c$ for $c \geq 0$ (this is of interest only when X, Y are correlated).

We derive our lower-bounds in terms of 2-dimensional regions, which can potentially yield stronger lower bounds than considering the two points $T_{\text{Wyn}}(XZ; YZ)$ and $T_{\text{Wyn}}(X; Y)$ on the one-dimensional line. The general relation between communication complexity and information complexity, and that between information complexity and tension ([Theorem 3](#) and [Theorem 1](#)) can be summarized as

$$\mathfrak{C} \subseteq \mathfrak{I} \subseteq \mathfrak{R},$$

where \mathfrak{C} denotes the set of communication cost pairs (number of bits from Alice to Bob, and vice-versa) achievable by protocols computing a possibly randomized function f , \mathfrak{I} denotes the information cost pairs (information communicated by Alice to Bob about her input, and vice versa) achievable by such protocols, and \mathfrak{R} , as described below, denotes a 2-dimensional restriction of the 3-dimensional “tension region” that was introduced

in [PP14]. Here, all three regions are defined to be “upward closed” subsets of \mathbb{R}_+^2 : i.e., if (x, y) is in the set and then so is (x', y') for all $x' \geq x$ and $y' \geq y$.

Before fully describing \mathfrak{R} , for simplicity, consider the case of independent X, Y . In this case, \mathfrak{R} is given by

$$\mathfrak{T}_0(XZ; YZ) = \{(r_1, r_2) \in \mathbb{R}_+^2 : \exists Q \text{ s.t. } XZ - Q - YZ \text{ and } I(XZ; Q|YZ) \leq r_1, I(YZ; Q|XZ) \leq r_2\}.$$

This is a convex, upward-closed region, typically bounded away from the origin. In the more general case, when X, Y are not independent, \mathfrak{R} is somewhat more complex. In particular, it is contained in the region

$$\mathfrak{T}_0(XZ; YZ) - \mathfrak{T}_0(X; Y) = \{(r_1, r_2) \in \mathbb{R}_+^2 : (r_1, r_2) + \mathfrak{T}_0(X; Y) \subseteq \mathfrak{T}_0(XZ; YZ)\}.$$

Typically, we expect the region $\mathfrak{T}_0(XZ; YZ)$ to be much further away from the origin than $\mathfrak{T}_0(X; Y)$ (i.e., $(XZ; YZ)$ has much higher tension than $(X; Y)$). The region $\mathfrak{T}_0(XZ; YZ) - \mathfrak{T}_0(X; Y)$ (or rather, the lower boundary of it) captures the least amount by which $\mathfrak{T}_0(X; Y)$ should be pushed away from the origin so that it moves completely inside $\mathfrak{T}_0(XZ; YZ)$. The bound $T_{\text{Wyn}}(XZ; YZ) - T_{\text{Wyn}}(X; Y)$ mentioned earlier, can be obtained as

$$\inf_{(a,b) \in \mathfrak{T}_0(XZ; YZ)} (a+b) - \inf_{(a,b) \in \mathfrak{T}_0(X; Y)} (a+b) \leq \inf_{(a,b) \in \mathfrak{T}_0(XZ; YZ) - \mathfrak{T}_0(X; Y)} (a+b).$$

Here we point out that the inequality above could be strict, in which case settling for a 1-dimensional version would give a weaker bound than what is implied by the 2-dimensional version.

The full definition of \mathfrak{R} is $\cap_{c \geq 0} \mathfrak{T}_c(XZ; YZ) - \mathfrak{T}_c(X; Y)$, where in $\mathfrak{T}_c(XZ; YZ)$ we do not restrict to Q such that $XZ - Q - YZ$; instead we require only that $I(XZ; YZ|Q) \leq c$. In showing that \mathfrak{R} gives a valid outer-bound on \mathfrak{J} , we rely on a certain “monotonicity” property of the 3-dimensional tension region of the views of the parties in a protocol: the tension region can only extend closer to the origin as the protocol progresses.²

While quite general in its form, we leave it as an open problem to exploit the full power of this connection, since understanding the full 3-dimensional tension region is an outstanding challenge.

Information Complexity vs. Communication Complexity. As mentioned above, the connection between information complexity and communication complexity is well-known. We extend this relation to the 2-dimensional regions \mathfrak{C} and \mathfrak{J} . Note that \mathfrak{C} corresponds to *average* communication-complexity. Hence $\mathfrak{C} \subseteq \mathfrak{J}$ directly yields a lower bounds not just on worst-case communication complexity (as it is often presented in the literature), but in fact on average communication complexity as well.³ This allows one to translate lower-bounds on information complexity of protocols of a certain error rate to lower-bounds on average communication complexity for the same error rate.

Discrepancy vs. Tension

Consider X, Y being n -bit long strings, and Z being a single bit with $\Pr[Z = f(X, Y)] \geq \frac{1}{2} + \epsilon$, where f is, say, the inner-product over $GF(2)$. When X, Y are independent, $T_{\text{Wyn}}(X; Y) = 0$. One would wonder if adding a single bit to the random variables can change their tension by more than a constant amount. But as it turns out, the correlation between XZ, YZ as captured by T_{Wyn} can be $\Omega(n)$ bits! For this, we rely on the function f having an exponentially small “discrepancy,” a combinatorial measure of complexity of a function.

Indeed, in [Section 5](#) we show that the Wyner-Tension $T_{\text{Wyn}}(XZ; YZ)$, where X, Y are independent, and $\Pr[Z = f(X, Y)] \geq \frac{1}{2} + \epsilon$, can be lower-bounded as $\Omega(\epsilon \log \frac{\epsilon}{\Delta})$ if the discrepancy of f (w.r.t. the distribution of (X, Y)) is upper-bounded by Δ . This compares favorably with a similar bound in [BW12], of the form $\Omega(\epsilon^2 \log \frac{\epsilon}{\Delta})$ (though, as mentioned above, the bound in [BW12] applies even if X, Y are not independent).

²A more general monotonicity property holds, allowing the parties to not just exchange messages, but also to “securely” delete parts of their views. This was shown in [PP14] for all of the tension region, including T_{Wyn} ; a similar result appeared for T_{GK} and two other points in the tension region in an earlier work of Wolf and Wullschlegel [WW05].

³In fact, we observe that the inequality $IC_\mu(\Pi) \leq CC(\Pi)$ [BR11] used to relate information cost and worst-case communication cost of a protocol can in fact be strengthened to $IC_\mu(\Pi) \leq CC_\mu(\Pi) \leq CC(\Pi)$, for any distribution μ over the inputs. (See [Lemma 1](#).)

To lower-bound $T_{\text{Wyn}}(XZ; YZ)$ it turns out to be enough to lower-bound $I(XY; Q)$ such that $X - Q - Y$ and given Q , Z is determined (i.e., $H(Z|Q) = 0$). The high-level intuition is to analyze the advantage Z has (i.e., $\Pr[Z = f(X, Y)] - \frac{1}{2}$) as contributed by different values of Q . For starters, suppose the input distribution is uniform and further, for each value q for Q , the conditional distribution $p_{XY|Q=q}$ is also *uniform over a rectangle*. Then, for q such that this rectangle is large, its contribution to the advantage will be small, because otherwise it will result in a large discrepancy (recall that Z must take a single value conditioned on $Q = q$). Thus, to achieve a large advantage when the discrepancy is small, most of the mass on Q should correspond to q such that $p_{XY|Q=q}$ is uniform over a “small” rectangle. Intuitively, this should imply a large value for $I(XY; Q)$.

This idea runs into several complications. Mainly, $p_{XY|Q=q}$ is guaranteed only to be a product distribution, and not necessarily uniform over its support. To tackle this, we show how to *slice* this distribution into several components, each of which is indeed uniform (or more generally, when XY is not uniform, each one is $p_{XY|(X,Y) \in r}$ for some rectangle r). One could then repeat the above argument with respect to the slices. However, including the index of the slice into Q would result in a large gap between its mutual information with XY , and that of the original Q . Instead we add a single bit to Q to indicate whether the slice is a large rectangle or a small rectangle. We then argue that collecting the small rectangles into one single subset will still result in a (relatively) small subset. With this, the above outline can indeed be made to work.

We remark that the intuition that if, for most q , the support of $p_{XY|Q=q}$ has a small mass in the original distribution p_{XY} , then $I(XY; Q)$ should be large is formalized in [Lemma 2](#). This may be of independent interest.

1.2 Related Work

Many of the recent advances in the field of communication complexity [[Yao79](#)] have followed from using various notions of information complexity. Earlier notions of information complexity appeared implicitly in several works [[Abi96](#), [PRV01](#), [SS02](#)], and was first explicitly defined in [[CSWY01](#)]. The current notion of (internal) information complexity originated in [[BYJKS04](#)]. Information complexity has been extensively used in the recent communication complexity literature [[BR11](#), [Bra12](#), [BW12](#), [CKW12](#), [KLL⁺12](#), [BBCR13](#)]. The notion was also adapted to specialized models or tasks [[JKS03](#), [JRS03](#), [JRS05](#), [HJMR10](#)]. The result in [[BW12](#)] (since generalized by [[KLL⁺12](#)]) relates most to the result we derive to illustrate the potential of tension bounds.

The notion of common information, to which tension is closely related, was developed in the information-theory literature [[GK73](#), [Wyn75](#), [AK74](#), [PP14](#)]. Recently, it has found use in communication complexity, cryptography and other problems in theoretical computer science, e.g. [[HJMR10](#), [BP13](#), [BJLP13](#), [DPP14](#)]. Some special cases of tension were implicit in the work of Wolf and Wullschleger [[WW05](#)], who used their monotonicity properties in a protocol to lower-bound the number of oblivious transfers needed for various secure computation tasks. The full-fledged notion of tension region was developed in [[PP14](#)]. A multi-party notion of tension was defined in [[PP12](#)].

2 Preliminaries

Notation. For brevity of notation, we shall often denote the random-variables (X, Y) etc. by XY etc. Also, we shall often use a random variable to denote the probability distribution of the random variable, when the random variables that it is jointly distributed with are clear from the context: i.e., we may write Q instead of $\mathbf{p}_{Q|XY}$. We write $A - Q - B$ to indicate that $I(A; B|Q) = 0$.

Communication Complexity. Let $\Pi(X; Y)$ be a (randomized) 2-party protocol with inputs to the two parties being X and Y respectively. The two parties alternate sending messages to each other; Π specifies which party sends the first message, and the function mapping each party’s current view to the distribution over the next message that it sends, and a distribution over an optional output it produces (on producing an output, the party

halts). The messages can be of arbitrary length, but should be self-terminating given the transcript so far, and either of the two inputs.⁴ For simplicity, we do not include public coins in our model; however, with suitable modifications in the definitions, all our results would continue to hold in such a model. In particular, we note that tension between two random variables is not altered by adding a common random variable (i.e., the public random tape) to both the random variables.

We write $\Pi(X; Y) \mapsto (A; B)$ to denote that the random variables $(A; B)$ (jointly distributed with $(X; Y)$) are the outputs produced by the two parties on running $\Pi(X; Y)$. We denote by $CC_{XY}^{(12)}(\Pi)$ (respectively, $CC_{XY}^{(21)}(\Pi)$) the expected number of bits sent by party 1 to party 2 (respectively, by party 2 to party 1) in the protocol $\Pi(X; Y)$; the expectation is over the randomness of the protocol, as well as the input distribution \mathbf{p}_{XY} .

The communication complexity – or more precisely, the “achievable communication rate region” – for computing $(A; B)$ given $(X; Y)$, is defined as:

$$\mathfrak{C}(A; B : X; Y) = \{(r_1, r_2) \in \mathbb{R}_+^2 : \exists \Pi \text{ s.t. } \Pi(X; Y) \mapsto (A; B) \text{ and } CC_{XY}^{(12)}(\Pi) \leq r_1, CC_{XY}^{(21)}(\Pi) \leq r_2\}.$$

Note that the region $\mathfrak{C}(A; B : X; Y)$ is an *upward closed region*. In fact, the different regions we shall define and use are all upward closed.

A special case of interest is when the $A = B = f(X, Y)$, for a boolean function $f : \mathcal{X} \times \mathcal{Y} \rightarrow \{0, 1\}$. In this case we shall typically require of a protocol that the two parties agree on the outcome, but we shall allow the outcome to be wrong with some probability ϵ (probability taken over the input distribution as well as the randomness of the protocol). We define the *communication complexity region* for f (for an error probability ϵ) to be:

$$\mathfrak{C}_\epsilon(f : X; Y) = \bigcup_{\substack{\mathbf{p}_{Z|XY}: \\ \text{SD}(\mathbf{p}_{ZXY}, \mathbf{p}_{f(X,Y)XY}) \leq \epsilon}} \mathfrak{C}(Z; Z : X; Y),$$

where $\text{SD}(p_A, p_B)$ is the total variation distance between the distributions p_A, p_B defined as $\text{SD}(p_A, p_B) = \frac{1}{2} \sum_a |p_A(a) - p_B(a)|$. Also of special interest is the (average-case) *communication complexity*, which considers just the total number of bits communicated, irrespective of the direction:

$$CC_{XY}^\epsilon(f) = \inf \{r_1 + r_2 : (r_1, r_2) \in \mathfrak{C}_\epsilon(f : X; Y)\}.$$

Information Complexity. The *information cost* of a protocol Π is defined as follows. Let $\Pi(X; Y) \mapsto (A; B)$ and let M denote the transcript of $\Pi(X; Y)$. Then we define

$$IC_{XY}^{(12)}(\Pi) = I(X; M|Y), \quad IC_{XY}^{(21)}(\Pi) = I(Y; M|X).$$

Then, $IC_{XY}(\Pi) = IC_{XY}^{(12)}(\Pi) + IC_{XY}^{(21)}(\Pi)$. We define the *information complexity region* as:

$$\mathfrak{I}(A; B : X; Y) = \{(r_1, r_2) \in \mathbb{R}_+^2 : \exists \Pi \text{ s.t. } \Pi(X; Y) \mapsto (A; B) \text{ and } IC_{XY}^{(12)}(\Pi) \leq r_1, IC_{XY}^{(21)}(\Pi) \leq r_2\}.$$

Of special interest is the following quantity — the information complexity of computing Z from $(X; Y)$.

$$IC_{XY}(Z) = \inf \{r_1 + r_2 : (r_1, r_2) \in \mathfrak{I}(Z; Z : X; Y)\}.$$

Discrepancy. Let $\mathcal{R} = \{\mathcal{X}' \times \mathcal{Y}' : \mathcal{X}' \subseteq \mathcal{X}, \mathcal{Y}' \subseteq \mathcal{Y}\}$, the set of all “rectangles” in $\mathcal{X} \times \mathcal{Y}$. Then, given a distribution \mathbf{p}_{XY} over $\mathcal{X} \times \mathcal{Y}$, and a boolean function $f : \mathcal{X} \times \mathcal{Y} \rightarrow \{0, 1\}$, we define

$$\begin{aligned} \text{Disc}_{XY}(f) &= \max_{r \in \mathcal{R}} |\Pr[(X, Y) \in r \wedge f(X, Y) = 0] - \Pr[(X, Y) \in r \wedge f(X, Y) = 1]| \\ &= \max_{\substack{\mathcal{X}' \subseteq \mathcal{X}, \\ \mathcal{Y}' \subseteq \mathcal{Y}}} \left| \sum_{\substack{(x,y) \in \mathcal{X}' \times \mathcal{Y}': \\ f(x,y)=0}} \mathbf{p}_{XY}(x, y) - \sum_{\substack{(x,y) \in \mathcal{X}' \times \mathcal{Y}': \\ f(x,y)=1}} \mathbf{p}_{XY}(x, y) \right|. \end{aligned}$$

⁴The traditional definition of a protocol in the communication complexity literature is slightly more restrictive: it requires that the messages are self-truncating, given just the transcript so far. We note that when the two parties have correlated inputs (e.g., as part of their private inputs, they share a one-time pad which is used to mask the entire communication) this should no more be required.

2.1 Tension

The tension region of a pair of random variables was defined in [PP14] as the following upward closed region.

Definition 1. For a pair of random variables A, B , their tension region $\mathfrak{T}(A; B)$ is defined as

$$\mathfrak{T}(A; B) = \{(r_1, r_2, r_3) : \exists Q \text{ jointly distributed with } A, B \\ \text{s.t. } I(B; Q|A) \leq r_1, I(A; Q|B) \leq r_2, I(A; B|Q) \leq r_3\}.$$

As shown in [PP14], without loss of generality, we may assume a cardinality bound $|Q| \leq |\mathcal{A}||\mathcal{B}| + 2$ on the alphabet Q in the above definition, where \mathcal{A} and \mathcal{B} are the alphabets of A and B , respectively. It was also shown there that $\mathfrak{T}(A; B)$ has the interpretation as a rate-information tradeoff region for a distributed common randomness generation problem which generalizes the common randomness problem of Gács and Körner [GK73]. $\mathfrak{T}(A; B)$ is a closed, convex region, with the following monotonicity property for randomized (public/private coins) protocols: Suppose X, Y are the inputs and A, B the outputs of the parties under a protocol. Let M denote the transcript of the protocol. Let $V_A = (X, A, M)$ and $V_B = (Y, B, M)$ denote the views of the parties at the end of the protocol.

Proposition 1 (Theorem 5.4 of [PP14]). $\mathfrak{T}(V_A; V_B) \supseteq \mathfrak{T}(X; Y)$.

In the sequel we will apply certain implications of the above result. Specifically, we will be interested in the inclusion relationship of certain restrictions of the tension regions of inputs and the views. For convenience, we define for $c \geq 0$ the intersection of tension region with the plane $r_3 = c$ as \mathfrak{T}_c . More precisely,

$$\mathfrak{T}_c(A; B) = \{(r_1, r_2) \in \mathbb{R}_+^2 : (r_1, r_2, c) \in \mathfrak{T}(A; B)\} \\ = \{(r_1, r_2) \in \mathbb{R}_+^2 : \exists \mathbf{p}_{Q|A,B} \text{ s.t. } I(B; Q|A) \leq r_1, I(A; Q|B) \leq r_2, I(A; B|Q) \leq c\}.$$

The case of $c = 0$ will be of special interest to us. Here, we will focus on the minimum $r_1 + r_2$. We define the Wyner-tension $T_{\text{Wyn}}(A; B)$ of two jointly distributed random variables A, B as

$$T_{\text{Wyn}}(A; B) = \inf\{r_1 + r_2 : (r_1, r_2) \in \mathfrak{T}_0(A; B)\} = \inf_{\substack{\mathbf{p}_{Q|AB}: \\ A-Q-B}} I(A; Q|B) + I(B; Q|A).$$

This quantity is related to Wyner's common information $CI_{\text{Wyn}}(A; B)$ of two random variables A, B [Wyn75].

$$CI_{\text{Wyn}}(A; B) = \inf_{\substack{\mathbf{p}_{Q|AB}: \\ A-Q-B}} I(A, B; Q).$$

It is easy to see the following [PP14].

$$T_{\text{Wyn}}(A; B) = CI_{\text{Wyn}}(A; B) - I(A; B).$$

Notice that $CI_{\text{Wyn}}(A; B) \geq I(A; B)$ and $T_{\text{Wyn}}(A; B) \geq 0$.

3 Tension vs. Information Complexity

In this section, we lower-bound information complexity in terms of tension. As we shall work with the more general information complexity region $\mathfrak{I}(A; B : X; Y)$, the “lower-bound” corresponds to bounding the region away from the origin. For this, we shall define a region $\mathfrak{R}(A; B : X; Y) \subseteq \mathbb{R}_+^2$, which will then be used to outer-bound the region $\mathfrak{I}(A; B : X; Y)$. We define:

$$\mathfrak{R}(A; B : X; Y) = \bigcap_{c \geq 0} (\mathfrak{T}_c(B, Y; A, X) - \mathfrak{T}_c(Y; X)),$$

where $S_1 - S_2 = \{(a, b) \in \mathbb{R}_+^2 : (a, b) + S_2 \subseteq S_1\}$ and $(a, b) + S$, for $a, b \in \mathbb{R}$ and $S \subseteq \mathbb{R}^2$, is $\{(x, y) \in \mathbb{R}^2 : (x + a, y + b) \in S\}$. We also define

$$\tilde{\mathfrak{R}}(A; B : X; Y) = (H(B|Y) - H(AB|XY), H(A|X) - H(AB|XY)) + \mathfrak{R}(A; B : X; Y).$$

Note that if $H(A|X) \geq H(AB|XY)$ and $H(B|Y) \geq H(AB|XY)$, then $\tilde{\mathfrak{R}}(A; B : X; Y) \subseteq \mathfrak{R}(A; B : X; Y)$. These conditions are satisfied if, for instance, $A = B$ (both parties output the same value), or $H(A, B|X, Y) = 0$ (the output is a deterministic function of the input), or more generally if $H(A|B, X, Y) = H(B|A, X, Y) = 0$ (i.e., any randomness in the outputs given the inputs is common to both outputs). Even if these conditions are not satisfied, if the outputs A and B are short, then $\tilde{\mathfrak{R}}(A; B : X; Y)$ is close to $\mathfrak{R}(A; B : X; Y)$, and the difference between the two can be ignored.

Theorem 1. $\mathfrak{I}(A; B : X; Y) \subseteq \tilde{\mathfrak{R}}(A; B : X; Y)$. In particular, if $H(A|X) \geq H(A, B|X, Y)$ and $H(B|Y) \geq H(A, B|X, Y)$, then,

$$\mathfrak{I}(A; B : X; Y) \subseteq \mathfrak{R}(A; B : X; Y).$$

Proof. Consider any protocol Π that takes $(X; Y)$ as input and outputs $(A; B)$. Let $U_A = (X, A)$, $U_B = (Y, B)$, the input-output of Alice and Bob; and let M be the transcript of the messages exchanged between Alice and Bob.

$$IC_{XY}^{(12)}(\Pi) = I(X; M|Y) \tag{1}$$

$$\begin{aligned} &\stackrel{(a)}{=} I(X; M, B|Y) = I(X; B|Y) + I(X; M|Y, B) \\ &= I(X; B|Y) - I(A; M|X, Y, B) + I(X, A; M|Y, B) \\ &\geq I(X; B|Y) - H(A|X, Y, B) + I(X, A; M|Y, B) \\ &= H(B|Y) - H(A, B|X, Y) + I(U_A; M|U_B), \end{aligned} \tag{2}$$

where (a) follows from the Markov chain $B - (Y, M) - X$. Similarly,

$$IC_{XY}^{(12)}(\Pi) \geq H(A|X) - H(A, B|X, Y) + I(U_B; M|U_A). \tag{3}$$

Then it is enough to outer bound the region containing $(I(U_A; M|U_B), I(U_B; M|U_A))$. Let $V_A = (U_A, M)$, $V_B = (U_B, M)$, the views of Alice and Bob at the end of the protocol. By **Proposition 1**,

$$\mathfrak{I}(V_B; V_A) \supseteq \mathfrak{I}(Y; X).$$

This implies that, for each $\mathbf{p}_{Q|X, Y}$, there exists a $\mathbf{p}_{\tilde{Q}|V_A, V_B}$ such that,

$$I(V_A; \tilde{Q}|V_B) \leq I(X; Q|Y), \tag{4}$$

$$I(V_B; \tilde{Q}|V_A) \leq I(Y; Q|X), \tag{5}$$

$$I(V_A; V_B|\tilde{Q}) \leq I(X; Y|Q). \tag{6}$$

But,

$$\begin{aligned} I(V_A; \tilde{Q}|V_B) &= I(U_A, M; \tilde{Q}|U_B, M) = I(U_A; \tilde{Q}, M|U_B) - I(U_A; M|U_B) \\ &\geq I(U_A; \tilde{Q}|U_B) - I(U_A; M|U_B) \end{aligned}$$

Similarly,

$$\begin{aligned} I(V_B; \tilde{Q}|V_A) &\geq I(U_B; \tilde{Q}|U_A) - I(U_B; M|U_A), \\ I(V_B; V_A|\tilde{Q}) &\geq I(U_A; U_B|\tilde{Q}). \end{aligned}$$

Using these in (4)-(6), we have that for all Q , there exists \tilde{Q} such that

$$\begin{aligned} I(U_A; M|U_B) + I(X; Q|Y) &\geq I(U_A; \tilde{Q}|U_B), \\ I(U_B; M|U_A) + I(Y; Q|X) &\geq I(U_B; \tilde{Q}|U_A), \\ I(X; Y|Q) &\geq I(U_A; U_B|\tilde{Q}). \end{aligned}$$

Hence, for every $c \geq 0$, we have

$$(I(U_A; M|U_B), I(U_B; M|U_A)) + \mathfrak{T}_c(Y; X) \subseteq \mathfrak{T}_c(U_B; U_A).$$

In other words, $(I(U_A; M|U_B), I(U_B; M|U_A))$ must lie in the set $\mathfrak{T}_c(U_B; U_A) - \mathfrak{T}_c(Y; X)$. Combined with (2) and (3), we get that $(IC_{XY}^{(12)}(\Pi), IC_{XY}^{(21)}(\Pi)) \in (H(B|Y) - H(AB|XY), H(A|X) - H(AB|XY)) + \bigcap_{c \geq 0} \mathfrak{T}_c(U_B; U_A) - \mathfrak{T}_c(Y; X)$. Since this holds for all Π such that $\Pi(X; Y) \mapsto (A; B)$, we get

$$\mathfrak{J}(A; B : X; Y) \subseteq (H(B|Y) - H(AB|XY), H(A|X) - H(AB|XY)) + \mathfrak{R}(A; B : X; Y) = \tilde{\mathfrak{R}}(A; B : X; Y).$$

Corollary 2. For all X, Y, Z ,

$$IC_{XY}(Z) \geq T_{\text{Wyn}}(XZ; YZ) - T_{\text{Wyn}}(X; Y).$$

In particular, if X and Y are independent of each other, $IC_{XY}(Z) \geq T_{\text{Wyn}}(XZ; YZ)$.

Proof. Firstly, note that the condition in **Theorem 1** holds when $A = B = Z$, since $H(Z|X) \leq H(Z|XY)$ and $H(Z|Y) \leq H(Z|XY)$. Thus,

$$\mathfrak{J}(Z; Z : X; Y) \subseteq \mathfrak{R}(Z; Z : X; Y) \subseteq \mathfrak{T}_0(YZ; XZ) - \mathfrak{T}_0(Y; X).$$

Then, $IC_{XY}(Z) = \inf_{(a,b) \in \mathfrak{J}(Z; Z : X; Y)} (a + b) \geq \inf_{(a,b) \in \mathfrak{T}_0(YZ; XZ) - \mathfrak{T}_0(Y; X)} (a + b)$. Now, $\forall (a, b) \in (S_1 - S_2)$, we have $S_1 \supseteq (a, b) + S_2$; hence,

$$\inf_{(r_1, r_2) \in S_1} (r_1 + r_2) \leq \inf_{(a, b) \in S_1 - S_2} (a + b) + \inf_{(r_1, r_2) \in S_2} (r_1 + r_2).$$

Recall that $\inf_{(r_1, r_2) \in \mathfrak{T}_0 U; V} (r_1 + r_2) = T_{\text{Wyn}}(U; V)$. Thus,

$$IC_{XY}(Z) \geq T_{\text{Wyn}}(YZ; XZ) - T_{\text{Wyn}}(Y; X).$$

The statement in the theorem follows from the symmetry of T_{Wyn} . \square

4 Information Complexity vs. Communication Complexity

Below we show that the communication complexity region is outer-bounded by the information complexity region. We start with **Lemma 1** below, which relates the communication cost pair of a protocol to its information cost pair. A simplified version of this result that has been used extensively, namely, $IC_{XY}(\Pi) \leq CC(\Pi)$, appears in [BR11]. Note that from **Lemma 1** it follows that, in fact, $IC_{XY}(\Pi) \leq CC_{XY}(\Pi)$ (and clearly, $CC_{XY}(\Pi) \leq CC(\Pi)$). That is, the information-complexity lower-bound applies not just to the worst case communication complexity, but also to the average case communication complexity.

Lemma 1. For any protocol Π and input distribution (X, Y) , the following hold:

$$IC_{XY}^{(12)}(\Pi) \leq CC_{XY}^{(12)}(\Pi), \quad IC_{XY}^{(21)}(\Pi) \leq CC_{XY}^{(21)}(\Pi).$$

In particular, $IC_{XY}(\Pi) \leq CC_{XY}(\Pi)$.

Proof. We shall show that $IC_{XY}^{(12)}(\Pi) \leq CC_{XY}^{(12)}(\Pi)$; the second inequality follows similarly, and the third is obtained by adding the first two inequalities. Below, the random variable M denotes the transcript of the protocol Π with input $(X; Y)$, M_i denotes the i^{th} bit of M , and M^i denotes the first i bits of M . For notational convenience, we define M_i to be a fixed symbol (say, 0) if i is greater than the length of M . Let \mathcal{M} be the set of all complete transcripts.⁵ Also, for $m \in \mathcal{M}$, we write $|m|_{12}$ to denote the (expected) number of bits in m that are sent by party 1 to party 2 (expectation over either input), and similarly $|m|_{21}$ to denote the bits in the other direction, so that $|m| = |m|_{12} + |m|_{21}$.

$$\begin{aligned}
IC_{XY}^{(12)}(\Pi) &= I(M; X|Y) = \sum_{i=0}^{\infty} I(M_{i+1}; X|Y, M^i) \\
&= \sum_{i=0}^{\infty} \sum_{m \in \{0,1\}^i} \Pr[M^i = m] \cdot I(M_{i+1}; X|Y, M^i = m) \\
&= \sum_{i=0}^{\infty} \sum_{m \in \{0,1\}^i} \left(\sum_{\substack{\hat{m} \in \mathcal{M}: \\ m = \hat{m}^i}} \Pr[M = \hat{m}] \right) \cdot I(M_{i+1}; X|Y, M^i = m) \\
&= \sum_{i=0}^{\infty} \sum_{\hat{m} \in \mathcal{M}} \Pr[M = \hat{m}] I(M_{i+1}; X|Y, M^i = \hat{m}^i) \\
&= \sum_{\hat{m} \in \mathcal{M}} \Pr[M = \hat{m}] \cdot \sum_{i=0}^{|\hat{m}|-1} I(M_{i+1}; X|Y, M^i = \hat{m}^i) \\
&\stackrel{(a)}{\leq} \sum_{\hat{m} \in \mathcal{M}} \Pr[M = \hat{m}] \cdot |\hat{m}|_{12} = CC_{XY}^{(12)}(\Pi)
\end{aligned}$$

where inequality (a) follows from the fact that, for each value of y , $I(M_{i+1}; X|Y = y, M^i = \hat{m}^i) = 0$ if, after \hat{m}^i (and given $Y = y$), the next message is sent by Bob, and otherwise $I(M_{i+1}; X|Y = y, M^i = \hat{m}^i) \leq H(M_{i+1}) \leq 1$. \square

The following theorem is an immediate consequence of [Lemma 1](#).

Theorem 3. $\mathfrak{C}(A; B : X; Y) \subseteq \mathfrak{J}(A; B : X; Y)$.

Proof. Consider any protocol Π that takes $(X; Y)$ as input and outputs $(A; B)$. By [Lemma 1](#), $IC_{XY}^{(12)}(\Pi) \leq CC_{XY}^{(12)}(\Pi)$ and $IC_{XY}^{(21)}(\Pi) \leq CC_{XY}^{(21)}(\Pi)$. Thus, by definition of $\mathfrak{J}(A; B : X; Y)$, $(CC_{XY}^{(12)}(\Pi), CC_{XY}^{(21)}(\Pi)) \in \mathfrak{J}(A; B : X; Y)$. Since this holds for all Π such that $\Pi(X; Y) \mapsto (A; B)$, and $\mathfrak{J}(A; B : X; Y)$ is an upward closed region, the theorem follows. \square

Following the definitions, the above theorem yields the following lower-bound:

$$CC_{X;Y}^{\epsilon}(f) \geq \inf_{\substack{\mathbf{p}_{Z|XY}: \\ \text{SD}(\mathbf{p}_{ZXY}, \mathbf{p}_{f(X,Y)XY}) \leq \epsilon}} IC_{XY}(Z).$$

Combining this with [Corollary 2](#), we obtain the following lower-bound on (average-case) communication complexity.

⁵Since we do not require the transcripts to be parsable on their own without an input (see [Footnote 4](#)), strictly speaking, the set of complete transcripts is not well-defined. However, \mathcal{M} can be defined more loosely as, for instance, the set of all strings of length d , where d is an upperbound on the worst-case communication cost of the protocol, and the arguments in the proof continue to hold. In fact, even if this cost is unbounded, but as long as the average cost $CC_{XY}^{(12)}(\Pi)$ is bounded (otherwise the inequality is trivial to see), it is possible to extend the proof by considering $d \rightarrow \infty$.

Corollary 4. For all $\epsilon \geq 0$, $CC_{X,Y}^\epsilon(f) \geq \inf_{\mathbf{p}_{Z|XY}: \text{SD}(\mathbf{p}_{ZXY}, \mathbf{p}_{f(X,Y)XY}) \leq \epsilon} T_{\text{Wyn}}(XZ; YZ) - T_{\text{Wyn}}(X; Y).$

In particular, if (X, Y) are independent of each other, $CC_{X,Y}^\epsilon(f) \geq \inf_{\mathbf{p}_{Z|XY}: \text{SD}(\mathbf{p}_{ZXY}, \mathbf{p}_{f(X,Y)XY}) \leq \epsilon} T_{\text{Wyn}}(XZ; YZ).$

5 Bounding Tension Using Discrepancy

Theorem 5. Suppose (X, Y) are independent random variables over $\mathcal{X} \times \mathcal{Y}$, and $f : \mathcal{X} \times \mathcal{Y} \rightarrow \{0, 1\}$ is a function with $\text{Disc}_{XY}(f) \leq \Delta$. Also, suppose Z is a binary random variable jointly distributed with (X, Y) such that $\Pr[Z \neq f(X, Y)] \leq \frac{1}{2} - \epsilon$. Then

$$T_{\text{Wyn}}(XZ; YZ) \geq \frac{\epsilon}{2(1-\epsilon)} \log \frac{\epsilon}{\Delta} - 4.$$

Proof. We seek to lower-bound the tension, $T_{\text{Wyn}}(XZ; YZ) = \inf_{Q: XZ - Q - YZ} I(XZ; Q|YZ) + I(YZ; Q|XZ)$. Consider a random variable Q over an alphabet \mathcal{Q} , jointly distributed with (X, Y) , such that $XZ - Q - YZ$. Firstly, note that this implies $H(Z|Q) = 0$, and $I(X; Y|Q) = 0$ (since both these quantities are upper-bounded by $I(XZ; YZ|Q) = 0$). To lower-bound $I(XZ; Q|YZ) + I(YZ; Q|XZ)$, it is enough to lower-bound $I(XY; Q)$, as shown below:

$$\begin{aligned} I(XZ; Q|YZ) + I(YZ; Q|XZ) &= I(X; Q|YZ) + I(Y; Q|XZ) \\ &= I(XZ; Q|Y) - I(Z; Q|Y) + I(YZ; Q|X) - I(Z; Q|X) \\ &\geq I(X; Q|Y) - 1 + I(Y; Q|X) - 1 \\ &= (I(X; Q|Y) + I(Y; Q|X) + I(X; Y)) - I(X; Y) - 2 \\ &= (I(XY; Q) + I(X; Y|Q)) - I(X; Y) - 2 \\ &= I(XY; Q) - I(X; Y) - 2, \end{aligned}$$

where in the last step we used the fact that $I(X; Y|Q) = 0$. Since we are given that X and Y are independent, we have $I(XZ; Q|YZ) + I(YZ; Q|XZ) \geq I(XY; Q) - 2$.

For all $q \in \mathcal{Q}$, let $D(q) = |\Pr[f(X, Y) = 0|Q = q] - \Pr[f(X, Y) = 1|Q = q]|$.

$$\begin{aligned} 2\epsilon &\leq \Pr[Z = f(X, Y)] - \Pr[Z \neq f(X, Y)] \\ &= \sum_{q \in \mathcal{Q}} \Pr[Q = q] (\Pr[Z = f(X, Y)|Q = q] - \Pr[Z \neq f(X, Y)|Q = q]) \\ &\leq \sum_{q \in \mathcal{Q}} \Pr[Q = q] D(q), \end{aligned}$$

where in the last step we used the fact that $H(Z|Q) = 0$.

We shall define an auxiliary random variable R over all rectangles (i.e., with alphabet $\mathcal{R} = \{\mathcal{X}' \times \mathcal{Y}' : \mathcal{X}' \subseteq \mathcal{X}, \mathcal{Y}' \subseteq \mathcal{Y}\}$), jointly distributed with (X, Y, Q) , satisfying that the following conditions for each $q \in \mathcal{Q}$. Below, let $\mathcal{R}_0 \subseteq \mathcal{R}$ denote the set of “small” rectangles: i.e., $\mathcal{R}_0 = \{r \in \mathcal{R} : \Pr[(X, Y) \in r] < \alpha\}$, where α is a parameter to be set later. Also, for $q \in \mathcal{Q}$, let $\mathcal{L}_q \subseteq \mathcal{X} \times \mathcal{Y}$ denote the set of all (x, y) which lie in the small rectangles that occur with q ; i.e.,

$$\mathcal{L}_q = \bigcup_{\substack{r \in \mathcal{R}_0: \\ \Pr[Q=q, R=r] > 0}} r.$$

Claim 1. There exists a random variable R with alphabet \mathcal{R} , jointly distributed with (X, Y, Q) such that for each $q \in \mathcal{Q}$ the following hold.

- For every $r \in \mathcal{R}$ such that $\Pr[Q = q, R = r] > 0$, the distribution $\mathbf{p}_{XY|Q=q, R=r}$ is the same as $\mathbf{p}_{XY|(X,Y) \in r}$ (i.e., \mathbf{p}_{XY} restricted to the rectangle r).
- $\Pr[(X, Y) \in \mathcal{L}_q] \leq 2\sqrt{\alpha}$.

We prove this claim in [Appendix B](#).

Let \hat{R} be a boolean random variable such that $\hat{R} = 0$ iff $R \in \mathcal{R}_0$, and $\hat{R} = 1$ otherwise. Let $Q' = (Q, \hat{R})$. Note that $I(XY; Q) \geq I(XY; Q') - 1$; so it is sufficient to lower-bound $I(XY; Q')$.

First, we lower-bound $\Pr[\hat{R} = 0]$, relying on the upper bound on discrepancy. Let $D(q, r) = |\Pr[f(X, Y) = 0|Q = q, R = r] - \Pr[f(X, Y) = 1|Q = q, R = r]|$. Then $D(q) \leq \sum_r \Pr[R = r|Q = q]D(q, r)$. Further,

$$\begin{aligned} \Pr[(X, Y) \in r] \cdot D(q, r) &= \Pr[(X, Y) \in r] \cdot |\Pr[f(X, Y) = 0|(X, Y) \in r] - \Pr[f(X, Y) = 1|(X, Y) \in r]| \\ &\quad \text{since } \mathbf{p}_{XY|Q=q, R=r} \equiv \mathbf{p}_{XY|(X,Y) \in r} \\ &= |\Pr[f(X, Y) = 0 \wedge (X, Y) \in r] - \Pr[f(X, Y) = 1 \wedge (X, Y) \in r]| \\ &\leq \text{Disc}_{XY}(f) \leq \Delta. \end{aligned}$$

Then, since $\Pr[(X, Y) \in r] \geq \alpha$ for $r \notin \mathcal{R}_0$, we conclude that $D(q, r) \leq \frac{\Delta}{\alpha}$, for $r \notin \mathcal{R}_0$. Now,

$$\begin{aligned} 2\epsilon &\leq \sum_{q \in \mathcal{Q}} \Pr[Q = q]D(q) \leq \sum_{q, r \in \mathcal{R}} \Pr[Q = q, R = r]D(q, r) \\ &\leq \sum_{q, r \in \mathcal{R}_0} \Pr[Q = q, R = r] + \sum_{q, r \notin \mathcal{R}_0} \Pr[Q = q, R = r]D(q, r) \\ &\leq \sum_{q, r \in \mathcal{R}_0} \Pr[Q = q, R = r] + \frac{\Delta}{\alpha} \sum_{q, r \notin \mathcal{R}_0} \Pr[Q = q, R = r] \\ &\leq \Pr[\hat{R} = 0] + \frac{\Delta}{\alpha}(1 - \Pr[\hat{R} = 0]). \end{aligned}$$

So, $\Pr[\hat{R} = 0] \geq \frac{2\epsilon - \frac{\Delta}{\alpha}}{1 - \frac{\Delta}{\alpha}}$.

Finally, we use the following lemma, proven in [Appendix B](#) (with $S = (X, Y)$, $T = Q'$ and $\mathcal{T}_0 = \mathcal{Q} \times \{0\}$) to obtain our lower bound on $I(XY; Q')$.

Lemma 2. *Let S, T be jointly distributed random variables over $\mathcal{S} \times \mathcal{T}$, and $\mathcal{T}_0 \subseteq \mathcal{T}$ be such that $\forall t \in \mathcal{T}_0$, $\Pr[S \in \mathcal{S}_t] \leq \delta$ where $\mathcal{S}_t = \{s \in \mathcal{S} : \Pr[S = s|T = t] > 0\}$, and $\Pr[T \in \mathcal{T}_0] \geq \varepsilon$. Then, $I(S; T) \geq \varepsilon \log \frac{1}{\delta}$.*

We apply this lemma with $\delta = 2\sqrt{\alpha}$ and $\varepsilon = \frac{2\epsilon - \frac{\Delta}{\alpha}}{1 - \frac{\Delta}{\alpha}}$. This yields $I(XY; Q') \geq \frac{2\epsilon - \frac{\Delta}{\alpha}}{1 - \frac{\Delta}{\alpha}}(\frac{1}{2} \log \frac{1}{\alpha} - 1)$. As described above, this bound on $I(XY; Q')$ yields the following bound on tension:

$$T_{\text{Wyn}}(XZ; YZ) \geq \frac{2\epsilon - \frac{\Delta}{\alpha}}{1 - \frac{\Delta}{\alpha}}(\frac{1}{2} \log \frac{1}{\alpha} - 1) - 3. \quad (7)$$

To complete the proof, we set $\alpha = \frac{\Delta}{\epsilon}$, and note that since $\epsilon < \frac{1}{2}$, we have $\epsilon/(1 - \epsilon) < 1$. \square

Remark: Often Δ is a quantity that vanishes as a size parameter of the inputs grows (e.g., when f is the inner-product function). When $\epsilon \cdot \log \frac{\epsilon}{\Delta} = \omega(1)$, one can obtain a tighter bound from the above proof, by setting $\alpha = (\frac{\Delta}{\epsilon})^{1-\beta}$ for a small enough $\beta > 0$. This gives $T_{\text{Wyn}}(XZ; YZ) \geq \epsilon \cdot \log \frac{\epsilon}{\Delta} \cdot (1 - o(1))$.

Acknowledgments

We gratefully acknowledge Mark Braverman, Prahladh Harsha and Rahul Jain for helpful discussions and pointers.

References

- [Abl96] Farid M. Ablayev. Lower bounds for one-way probabilistic communication complexity and their application to space complexity. *Theor. Comput. Sci.*, 157(2):139–159, 1996. 4
- [AK74] Rudolf Ahlswede and János Körner. On common information and related characteristics of correlated information sources. In *7th Prague Conference on Information Theory*, 1974. 4
- [BBCR13] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. *SIAM J. Comput.*, 42(3):1327–1363, 2013. 4
- [BJLP13] Gábor Braun, Rahul Jain, Troy Lee, and Sebastian Pokutta. Information-theoretic approximations of the nonnegative rank. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:158, 2013. 4
- [BP13] Gábor Braun and Sebastian Pokutta. Common information and unique disjointness. In *FOCS*, pages 688–697, 2013. 4
- [BR11] Mark Braverman and Anup Rao. Information equals amortized communication. In *FOCS*, pages 748–757, 2011. 3, 4, 8
- [Bra12] Mark Braverman. Interactive information complexity. In *STOC*, pages 505–524, 2012. 4
- [BW12] Mark Braverman and Omri Weinstein. A discrepancy lower bound for information complexity. In *APPROX-RANDOM*, pages 459–470, 2012. 1, 3, 4
- [BYJKS04] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004. 4
- [CK81] Imre Csiszár and János Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Akadémiai Kiadó, Budapest, 1981. 14
- [CKW12] Amit Chakrabarti, Ranganath Kondapally, and Zhenghui Wang. Information complexity versus corruption and applications to orthogonality and gap-hamming. In *APPROX-RANDOM*, pages 483–494, 2012. 4
- [CSWY01] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Chi-Chih Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *FOCS*, pages 270–278, 2001. 4
- [DPP14] Deepesh Data, Manoj M. Prabhakaran, and Vinod M. Prabhakaran. On the communication complexity of secure computation. In *Advances in Cryptology - CRYPTO 2014 - 34th Annual Cryptology Conference, Santa Barbara, CA, USA, August 17-21, 2014, Proceedings, Part II*, pages 199–216, 2014. 1, 4
- [GK73] Peter Gács and János Körner. Common information is far less than mutual information. *Problems of Control and Information Theory*, 2(2):149–162, 1973. 1, 4, 6, 13
- [HJMR10] Prahladh Harsha, Rahul Jain, David McAllester, and Jaikumar Radhakrishnan. The communication complexity of correlation. *IEEE Transactions on Information Theory*, 56(1):438–449, 2010. 4
- [JKS03] T. S. Jayram, Ravi Kumar, and D. Sivakumar. Two applications of information complexity. In *STOC*, pages 673–682, 2003. 4

- [JRS03] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A direct sum theorem in communication complexity via message compression. In *ICALP*, pages 300–315, 2003. 4
- [JRS05] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Prior entanglement, message compression and privacy in quantum communication. In *IEEE Conference on Computational Complexity*, pages 285–296, 2005. 4
- [KLL⁺12] Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. Lower bounds on information complexity via zero-communication protocols and applications. In *FOCS*, pages 500–509, 2012. 4
- [PP12] Manoj Prabhakaran and Vinod Prabhakaran. On secure multiparty sampling for more than two parties. In *Proceedings of the 2012 IEEE International Information Theory Workshop (ITW 2012)*, 2012. 4
- [PP14] Vinod M. Prabhakaran and Manoj M. Prabhakaran. Assisted common information with an application to secure two-party sampling. *IEEE Transactions on Information Theory*, 60(6):3413–3434, 2014. 1, 3, 4, 6, 13, 14, 15
- [PRV01] Stephen J Ponzio, Jaikumar Radhakrishnan, and Srinivasan Venkatesh. The communication complexity of pointer chasing. *Journal of Computer and System Sciences*, 62(2):323–355, 2001. 4
- [SS02] Michael E. Saks and Xiaodong Sun. Space lower bounds for distance approximation in the data stream model. In *STOC*, pages 360–369, 2002. 4
- [WW05] Stefan Wolf and Jürg Wullschleger. New monotones and lower bounds in unconditional two-party computation. In *CRYPTO*, pages 467–477, 2005. 3, 4, 15
- [Wyn75] Aaron D. Wyner. The common information of two dependent random variables. *IEEE Transactions on Information Theory*, 21(2):163–179, 1975. 1, 4, 6, 14
- [Yao79] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *STOC*, pages 209–213, 1979. 1, 4

A On The Nature of Tension Region

In this appendix we present a gentle introduction to the notion of tension region, as developed in [PP14]. We refer the interested readers to [PP14] for more details.

Consider the random variables $X = (X', Q)$ and $Y = (Y', Q)$ where X', Y', Q are independent. In this case, it is natural to consider Q as the common random variable of X and Y and $H(Q)$ as a natural measure of “common information.” Q is determined both by X and by Y individually. Moreover, conditioned on Q , X and Y are independent, i.e., $X - Q - Y$ is a Markov chain. One could extend this to arbitrary X, Y , in a couple of natural ways. The approach of Gács and Körner [GK73] is to find the “largest” random variable Q (largeness being measured in terms of entropy) such that it is determined by X alone as well as by Y alone (with probability 1):

$$\begin{aligned}
 CI_{\text{GK}}(X; Y) &= \max_{\substack{\mathbf{p}_{Q|XY}: \\ H(Q|X)=H(Q|Y)=0}} H(Q) \\
 &= I(X; Y) - \min_{\substack{\mathbf{p}_{Q|XY}: \\ H(Q|X)=H(Q|Y)=0}} I(X; Y|Q).
 \end{aligned}$$

Clearly $CI_{\text{GK}}(X; Y) \leq I(X; Y)$ and, in general, this inequality may be strict, i.e., common information, in general, does not account for all the dependence between X and Y .

Wyner gave a different generalization [Wyn75] where he defined common information in terms of the “smallest” random variable Q (smallness being measured in terms of $I(XY; Q)$) so that X and Y are independent conditioned on Q .

$$\begin{aligned} CI_{\text{Wyn}}(X; Y) &= \min_{\substack{\mathbf{p}_{Q|XY}: \\ X-Q-Y}} I(XY; Q) \\ &= I(X; Y) + \min_{\substack{\mathbf{p}_{Q|XY}: \\ X-Q-Y}} (I(Y; Q|X) + I(X; Q|Y)). \end{aligned}$$

Now $CI_{\text{Wyn}}(X; Y) \geq I(X; Y)$. When X, Y are of the form $X = (X', Q)$ and $Y = (Y', Q)$, where X', Y', Q are independent, then there indeed is a unique interpretation of common information (when $CI_{\text{GK}}(X; Y) = CI_{\text{Wyn}}(X; Y) = H(Q)$). Between these extremes represented by these two measures, there are several ways in which one could define a random variable to capture the dependence between X and Y .

Definition 2. For a pair of correlated random variables (X, Y) , and $\mathbf{p}_{Q|XY}$, we say Q perfectly resolves (X, Y) if $I(X; Y|Q) = 0$ and $H(Q|X) = H(Q|Y) = 0$. We say (X, Y) is perfectly resolvable if there exists $\mathbf{p}_{Q|XY}$ such that Q perfectly resolves (X, Y) .

If (X, Y) is perfectly resolvable, then $CI_{\text{GK}}(X; Y) = I(X; Y) = CI_{\text{Wyn}}(X; Y)$ represents the entire mutual information between them. Tension region $\mathfrak{T}(X; Y)$ can be thought of as measuring the extent to which a pair of random variables (X, Y) is *not* resolvable.

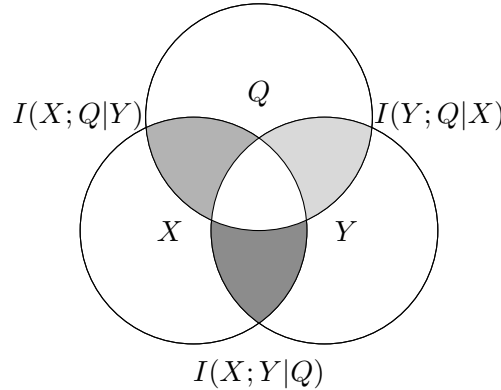


Figure 1 A Venn diagram representation of the three coordinates $(I(Y; Q|X), I(X; Q|Y), I(X; Y|Q))$ in the definition of $\mathfrak{T}(X; Y)Q$. Figure taken from [PP14].

Recall the definition of *tension region* $\mathfrak{T}(A; B)$ of a pair of random variables A, B :

$$\begin{aligned} \mathfrak{T}(A; B) &= \{(r_1, r_2, r_3) : \exists Q \text{ jointly distributed with } A, B \\ &\quad \text{s.t. } I(B; Q|A) \leq r_1, I(A; Q|B) \leq r_2, I(A; B|Q) \leq r_3\}. \end{aligned}$$

It follows from Fenchel-Eggleston’s strengthening of Carathéodory’s theorem [CK81, pg. 310], that we can restrict ourselves to $\mathbf{p}_{Q|XY}$ with alphabet \mathcal{Q} such that $|\mathcal{Q}| \leq |\mathcal{X}||\mathcal{Y}| + 2$.

It can be shown that $\mathfrak{T}(X; Y)$ includes the origin if and only if the pair (X, Y) is perfectly resolvable. When this is not the case, it is important to consider all three coordinates of together to identify the unresolvable nature of a pair (X, Y) , because since $\mathfrak{T}(X; Y)$ does intersect each of the three axes, or in other words, any two coordinates of can be made simultaneously 0 by choosing an appropriate Q .

Below we summarize several useful properties of $\mathfrak{T}(X; Y)$. For interpretations of $\mathfrak{T}(X; Y)$ in terms of certain information theoretic problems, we refer the reader to [PP14].

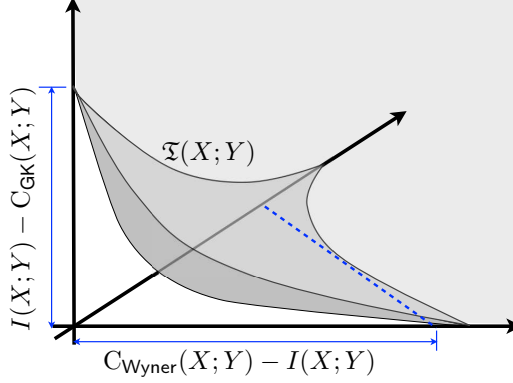


Figure 2 A schematic representation of the region $\mathfrak{T}(X; Y)$. $\mathfrak{T}(X; Y)$ is an unbounded, convex region, bounded away from the origin (unless (X, Y) is perfectly resolvable). Relationship between two points on the boundary of $\mathfrak{T}(X; Y)$ and the quantities $C_{GK}(X; Y)$ and $C_{Wyn}(X; Y)$ (The dotted line is at 45° to the axes.) Figure taken from [PP14].

A.1 Some Properties of Tension

Monotonicity of $\mathfrak{T}(X; Y)$. Wolf and Wullschlegel [WW05] showed that the three axes intercepts have a certain “monotonicity” property (they can only decrease, as X, Y evolve as the views of two parties in a protocol). In fact, this monotonicity is a consequence of the monotonicity of the entire region $\mathfrak{T}(X; Y)$ stated in [Proposition 1](#).

Tensorization of $\mathfrak{T}(X; Y)$. If (X_1, Y_1) is independent of (X_2, Y_2) , then

$$\mathfrak{T}((X_1 X_2); (Y_1 Y_2)) = \mathfrak{T}(X_1; Y_1) + \mathfrak{T}(X_2; Y_2).$$

Convexity, closedness, and continuity of $\mathfrak{T}(X; Y)$. Firstly, the region of tension is closed and convex. Secondly, the region of tension is *continuous* in the sense that when the joint p.m.f. $p_{X,Y}$ is close to the joint p.m.f. $p_{X',Y'}$, the tension regions $\mathfrak{T}(X; Y)$ and $\mathfrak{T}(X'; Y')$ are also close. Specifically, if $\text{SD}(XY, X'Y') \leq \epsilon$, then $\mathfrak{T}(X; Y) \subseteq \mathfrak{T}(X'; Y') - \delta(\epsilon)$, where $\delta(\epsilon) = 2H_2(\epsilon) + \epsilon \log \max\{|\mathcal{X}|, |\mathcal{Y}|\}$.

B Proof of [Lemma 2](#) and [Claim 1](#).

To complete the proof of [Theorem 5](#) we need to prove [Lemma 2](#) and [Claim 1](#). We do this below.

Proof of [Lemma 2](#). We have

$$\begin{aligned} I(S; T) &= \sum_{(s,t) \in \mathcal{S} \times \mathcal{T}} \mathbf{p}_{S,T}(s,t) \log \frac{\mathbf{p}_{S,T}(s,t)}{\mathbf{p}_S(s) \mathbf{p}_T(t)} \\ &= \sum_{t \in \mathcal{T}} \mathbf{p}_T(t) \sum_{s \in \mathcal{S}_t} \mathbf{p}_{S|T}(s|t) \log \frac{\mathbf{p}_{S|T}(s|t)}{\mathbf{p}_S(s)} \\ &= \sum_{t \in \mathcal{T}_0} \mathbf{p}_T(t) \sum_{s \in \mathcal{S}_t} \mathbf{p}_{S|T}(s|t) \log \frac{\mathbf{p}_{S|T}(s|t)}{\mathbf{p}_S(s)} + \sum_{t \in \mathcal{T} - \mathcal{T}_0} \mathbf{p}_T(t) \sum_{s \in \mathcal{S}_t} \mathbf{p}_{S|T}(s|t) \log \frac{\mathbf{p}_{S|T}(s|t)}{\mathbf{p}_S(s)} \end{aligned}$$

Notice that, for each t

$$\sum_{s \in \mathcal{S}_t} \mathbf{p}_{S|T}(s|t) \log \frac{\mathbf{p}_{S|T}(s|t)}{\mathbf{p}_S(s)} = D(\mathbf{p}_{S|T=t} \| \mathbf{p}_S) \geq 0.$$

Hence, we have

$$I(S; T) \geq \sum_{t \in \mathcal{T}_0} \mathbf{p}_T(t) \sum_{s \in \mathcal{S}_t} \mathbf{p}_{S|T}(s|t) \log \frac{\mathbf{p}_{S|T}(s|t)}{\mathbf{p}_S(s)}.$$

For each $t \in \mathcal{T}_0$, let $p_t = \Pr[S \in \mathcal{S}_t] = \sum_{s \in \mathcal{S}_t} \mathbf{p}_S(s)$, and let us define over \mathcal{S}_t the probability mass function, $\mathbf{p}_{(t)}(s) = \frac{\mathbf{p}_S(s)}{p_t}$, $s \in \mathcal{S}_t$. Note that $p_t \leq \delta$. Then, for $t \in \mathcal{T}_0$,

$$\begin{aligned} \sum_{s \in \mathcal{S}_t} \mathbf{p}_{S|T}(s|t) \log \frac{\mathbf{p}_{S|T}(s|t)}{\mathbf{p}_S(s)} &= \sum_{s \in \mathcal{S}_t} \mathbf{p}_{S|T}(s|t) \log \frac{\mathbf{p}_{S|T}(s|t)}{\mathbf{p}_S(s)/p_t} \frac{1}{p_t} \\ &= D(\mathbf{p}_{S|T} \| \mathbf{p}_{(t)}) + \log \frac{1}{p_t} \\ &\geq \log \frac{1}{\delta}. \end{aligned}$$

Substituting this back,

$$I(S; T) \geq \sum_{t \in \mathcal{T}_0} \mathbf{p}_T(t) \log \frac{1}{\delta} \geq \varepsilon \log \frac{1}{\delta}. \quad \square$$

Proof of Claim 1. It remains to describe the distribution $\mathbf{p}_{R|XYQ}$ so that the conditions listed in Claim 1 hold.

For $r = \mathcal{X}_r \times \mathcal{Y}_r \in \mathcal{R}$, we let

$$\begin{aligned} \sigma_{q,r} &= \min_{x \in \mathcal{X}_r} \frac{\Pr[X = x, Q = q]}{\Pr[X = x] \Pr[Q = q]} - \max_{x' \notin \mathcal{X}_r} \frac{\Pr[X = x', Q = q]}{\Pr[X = x'] \Pr[Q = q]} \\ \tau_{q,r} &= \min_{y \in \mathcal{Y}_r} \frac{\Pr[Y = y, Q = q]}{\Pr[Y = y] \Pr[Q = q]} - \max_{y' \notin \mathcal{Y}_r} \frac{\Pr[Y = y', Q = q]}{\Pr[Y = y'] \Pr[Q = q]} \end{aligned}$$

Above, in defining $\max_{x' \notin \mathcal{X}_r}$, if no such x' exists – i.e., $\mathcal{X}_r = \mathcal{X}$ – we take the maximum to be 0 (and similarly for $\max_{y' \notin \mathcal{Y}_r}$). Now we define $\mathbf{p}_{R|XYQ}$ as follows:

$$\Pr[R = r | X = x, Y = y, Q = q] = \begin{cases} \sigma_{q,r} \cdot \tau_{q,r} \cdot \frac{\Pr[X=x, Y=y]}{\Pr[X=x, Y=y | Q=q]} & \text{if } \sigma_{q,r} > 0, \tau_{q,r} > 0 \text{ and } (x, y) \in r \\ 0 & \text{otherwise.} \end{cases}$$

An alternate way to describe the mass assigned to r is as follows. Let $\mathcal{X}_q \times \mathcal{Y}_q$ be the support of $\mathbf{p}_{XY|Q=q}$. Let $\mathcal{X}_q = \{x_1, \dots, x_M\}$, such that $\frac{\Pr[X=x_i, Q=q]}{\Pr[X=x_i] \Pr[Q=q]} \geq \frac{\Pr[X=x_{i+1}, Q=q]}{\Pr[X=x_{i+1}] \Pr[Q=q]}$ for all $i \in [1, M-1]$. For notational convenience, we also define a dummy x_{M+1} with $\frac{\Pr[X=x_{M+1}, Q=q]}{\Pr[X=x_{M+1}] \Pr[Q=q]} = 0$. Define y_1, \dots, y_N, y_{N+1} similarly, where $N = |\mathcal{Y}_q|$. Then, the only rectangles r for which $\Pr[R = r | Q = q]$ can be positive are of the form $r_{ij} = \mathcal{X}_i \times \mathcal{Y}_j$ for $(i, j) \in [M] \times [N]$, where $\mathcal{X}_i = \{x_1, \dots, x_i\}$, $\mathcal{Y}_j = \{y_1, \dots, y_j\}$, $\frac{\Pr[X=x_i, Q=q]}{\Pr[X=x_i] \Pr[Q=q]} > \frac{\Pr[X=x_{i+1}, Q=q]}{\Pr[X=x_{i+1}] \Pr[Q=q]}$, and $\frac{\Pr[Y=y_j, Q=q]}{\Pr[Y=y_j] \Pr[Q=q]} > \frac{\Pr[Y=y_{j+1}, Q=q]}{\Pr[Y=y_{j+1}] \Pr[Q=q]}$.

First, we verify that $\mathbf{p}_{R|Q=q, X=x, Y=y}$ is indeed a valid probability distribution.

$$\begin{aligned}
& \sum_{r \in \mathcal{R}} \Pr[R = r | Q = q, X = x_{i^*}, Y = y_{i^*}] \\
&= \sum_{r: (x_{i^*}, y_{i^*}) \in r} \sigma_{q,r} \cdot \tau_{q,r} \cdot \frac{\Pr[X = x_{i^*}, Y = y_{i^*}]}{\Pr[X = x_{i^*}, Y = y_{i^*} | Q = q]} \\
&= \frac{\Pr[X = x_{i^*}, Y = y_{i^*}]}{\Pr[X = x_{i^*}, Y = y_{i^*} | Q = q]} \cdot \sum_{i=i^*}^M \sum_{j=j^*}^N \sigma_{q,r_{ij}} \cdot \tau_{q,r_{ij}} \\
&= \frac{\Pr[X = x_{i^*}, Y = y_{i^*}]}{\Pr[X = x_{i^*}, Y = y_{i^*} | Q = q]} \cdot \sum_{i=i^*}^M \left(\frac{\Pr[X = x_i, Q = q]}{\Pr[X = x_i] \Pr[Q = q]} - \frac{\Pr[X = x_{i+1}, Q = q]}{\Pr[X = x_{i+1}] \Pr[Q = q]} \right) \\
&\quad \cdot \sum_{j=j^*}^N \left(\frac{\Pr[Y = y_j, Q = q]}{\Pr[Y = y_j] \Pr[Q = q]} - \frac{\Pr[Y = y_{j+1}, Q = q]}{\Pr[Y = y_{j+1}] \Pr[Q = q]} \right) \\
&= \frac{\Pr[X = x_{i^*}, Y = y_{i^*}]}{\Pr[X = x_{i^*}, Y = y_{i^*} | Q = q]} \cdot \frac{\Pr[X = x_{i^*}, Q = q]}{\Pr[X = x_{i^*}] \Pr[Q = q]} \cdot \frac{\Pr[Y = y_{j^*}, Q = q]}{\Pr[Y = y_{j^*}] \Pr[Q = q]} = 1,
\end{aligned}$$

where in the last step we used the facts that X, Y are independent and also they are conditionally independent conditioned on Q .

Next, we verify that $\mathbf{p}_{XY|Q=q, R=r} \equiv \mathbf{p}_{XY|(X,Y) \in r}$. Firstly, if $(x, y) \notin r$, then $\Pr[R = r | X = x, Y = y, Q = q] = 0$, and hence $\Pr[X = x, Y = y | Q = q, R = r] = 0$ (and also, $\Pr[X = x, Y = y | (X, Y) \in r] = 0$). Now, suppose $(x, y) \in r$. Then,

$$\begin{aligned}
\Pr[X = x, Y = y | Q = q, R = r] &= \frac{\Pr[R = r | X = x, Y = y, Q = q] \Pr[X = x, Y = y | Q = q]}{\Pr[R = r | Q = q]} \\
&= \frac{\sigma_{q,r} \cdot \tau_{q,r} \cdot \Pr[X = x, Y = y]}{\Pr[R = r | Q = q]} = \frac{\Pr[X = x, Y = y]}{F(q, r)},
\end{aligned}$$

where $F(q, r)$ is a quantity independent of (x, y) . Since $\Pr[X = x, Y = y | Q = q, R = r]$ is a probability distribution, $F(q, r) = \sum_{(x,y) \in r} \Pr[X = x, Y = y] = \Pr[(X, Y) \in r]$. Thus indeed, $\Pr[X = x, Y = y | Q = q, R = r] = \Pr[X = x, Y = y | (X, Y) \in r]$.

Finally, we argue that $\Pr[(X, Y) \in \mathcal{L}_q] \leq 2\sqrt{\alpha}$. Consider any $q \in \mathcal{Q}$, and as before, let $\mathcal{X}_q = \{x_1, \dots, x_M\}$, $\mathcal{Y}_q = \{y_1, \dots, y_N\}$ sorted appropriately, and, for $i \in [M], j \in [N]$, $r_{ij} = \{x_1, \dots, x_i\} \times \{y_1, \dots, y_j\}$. Then $(x, y) \in \mathcal{L}_q$ iff $(x, y) \in r_{ij}$ for some $r_{ij} \in \mathcal{R}_0$ (i.e., $\Pr[(X, Y) \in r_{ij}] \leq \alpha$). Let i^* be the maximum value in $[M]$ such that $\Pr[X \in \{x_1, \dots, x_{i^*}\}] \leq \sqrt{\alpha}$, and similarly, let j^* be the maximum value in $[N]$ such that $\Pr[Y \in \{y_1, \dots, y_{j^*}\}] \leq \sqrt{\alpha}$. Then we note that, if $i > i^*$ and $j > j^*$, then $(x_i, y_j) \notin \mathcal{L}_q$. This is because, $(x_i, y_j) \in r_{i'j'} \implies (i' \geq i > i^*, j' \geq j > j^*) \implies r_{i'j'} \notin \mathcal{R}_0$, as $\Pr[(X, Y) \in r_{i'j'}] = \Pr[X \in \{x_1, \dots, x_{i'}\}] \cdot \Pr[Y \in \{y_1, \dots, y_{j'}\}] > \sqrt{\alpha}\sqrt{\alpha}$ (by definition of i^* and j^*). Hence,

$$\begin{aligned}
\Pr[(X, Y) \in \mathcal{L}_q] &\leq \Pr[(X \in \{x_1, \dots, x_{i^*}\}) \vee (Y \in \{y_1, \dots, y_{j^*}\})] \\
&\leq \Pr[X \in \{x_1, \dots, x_{i^*}\}] + \Pr[Y \in \{y_1, \dots, y_{j^*}\}] \leq 2\sqrt{\alpha}.
\end{aligned}$$

□